ELSEVIER

Journal of Cystic Fibrosis 4 (2005) 193 – 196

Journal of Cystic Fibrosis

www.elsevier.com/locate/jcf

Short communication

Overrepresentation of immunostimulatory CpG motifs in *Burkholderia* genomes

Tom Coenye*, Peter Vandamme

Laboratorium voor Farmaceutische, Microbiologie, Universiteit Gent, Harelbekestraat 72, B-9000 Gent, Belgium

Received 17 January 2005; received in revised form 2 February 2005; accepted 23 March 2005
Available online 15 June 2005

Abstract

Pulmonary infections with *Burkholderia cepacia* complex organisms contribute significantly to morbidity and mortality in patients with cystic fibrosis (CF), partially due to the intense inflammatory response of the host to the presence of bacteria and their byproducts. In the present study we show that *Burkholderia* genomes contain a large number of immunostimulatory CpG motifs. This is mainly because of their large genome size. This suggests that DNA from *Burkholderia* sp. has the potential to cause significant inflammatory response. Whether this contributes significantly to the airway inflammation often observed in infected CF patients remains to be determined.

Keywords: *Burkholderia cepacia* complex; Cystic fibrosis; Immunostimulatory DNA; CpG motif

1. Introduction

In persons with cystic fibrosis (CF), pulmonary disease is the major cause of morbidity and mortality; mutations in the *cftr* gene result in an altered airway surface liquid (ASL) which in some way predisposes the lungs to bacterial infections [1]. Organisms typically associated with the respiratory tract of CF patients are *Staphylococcus aureus*, *Haemophilus influenzae*, *Pseudomonas aeruginosa* and *Burkholderia cepacia* complex bacteria [2]. Due to the high viscosity of respiratory secretions and the altered ASL, normal host defense mechanisms are impaired and despite the intense inflammatory response, the host does not have the ability to clear the infection [3]. The continuous presence of bacteria and their byproducts in the lungs of CF patients leads to a vicious cycle in which the host immune response leads to more tissue damage, leading to more infection and further inflammation [4]. There are strong indications that the inflammatory response to *P. aeruginosa* infection is induced by lipopysaccharide (LPS), phospholipase C and exotoxin A [5–7]. Similarly, several studies have shown that LPS from *B. cepacia* complex bacteria can activate neutrophils [8] and stimulate tumor necrosis factor α production [9,10]. There are also indications that during infections with *Burkholderia pseudomallei* and *Burkholderia mallei* (the causative agents of melioidosis and glanders, respectively) an intense inflammatory response is mounted (see for example Refs. [11,12]).

It has been known for over 20 years that bacterial DNA can activate the human immune system [13–15]. Several studies in the early 1990s showed that the immunostimulatory properties of bacterial DNA could be attributed to the presence of unmethylated CpG dinucleotides [16–18]. Bacterial DNA can induce B-cells to proliferate and secrete immunoglobulines [19], can induce the production of immunomodulatory cytokines [20], can activate the acute phase response [21] and can activate specific subsets of dendritic cells [22]. It is thought that the response to CpG motifs is mediated by the Toll-like receptor 9 (TLR9) [23–25] and differences in human and murine TLR9 specificity might explain why the immune system of both organisms is stimulated by different CpG-containing motifs [25–27]. Several studies have indicated that TLR9 mediates the immune response in a concentration-dependent way, i.e.

* Corresponding author. Tel.: +32 (9) 264 8093; fax: +32 (9) 264 8195.
E-mail address: Tom.Coenye@UGent.be (T. Coenye).

1569-1993/$ - see front matter © 2005 European Cystic Fibrosis Society. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jcf.2005.03.001

DNA molecules containing multiple CpG motifs have a greater stimulatory capacity [28,29]. In addition, it has been suggested that the immunomodulatory effects of DNA may augment the inflammatory response to endotoxin, indicating that the presence of bacterial DNA might increase the sensitivity to toxic effects of LPS. In contrast to bacterial DNA, vertebrate DNA shows a significant CpG suppression [30,31] and it has been postulated that the vertebrate immune system has evolved to recognise unmethylated CpG motifs and can respond to them with a rapid and coordinated cytokine response, leading to induction of humoral and cell-mediated immunity.

In the present study we compared the occurrence of CpG-containing motifs in the *Burkholderia cenocepacia* strain J2315 genome with the occurrences in other bacterial genomes, including those of other CF pathogens and other *Burkholderia* species.

2. Materials and methods

Complete genome sequences used in the present study are shown in Table 1. We included all completely sequenced *Burkholderia* genomes, as well as large contigs of other *Burkholderia* genomes. For comparison we also included (i) genome sequences of other CF pathogens and (ii) genome sequences of other organisms with a comparable GC content and/or genome size. To determine the dinucleotide relative abundance and absolute number and frequencies of CpG-containing motifs of different lengths, sequences were first concatenated with their inverted complementary sequence using the EMBOSS package (http://www.hgmp.mrc.ac.uk/software/EMBOSS). Dinucleotide relative abundances ρ^*_{XY} were calculated using the equation $\rho^*_{XY}=f_{XY}/f_Xf_Y$ where f_{XY} denotes the frequency of dinucleotide XY and f_X and f_Y denote the frequencies of X and Y, respectively [32]. Statistical theory and data from previous studies [31,32] indicate that the normal range of ρ^*_{XY} is between 0.78 and 1.23. Absolute number and frequencies of the CpG-containing motifs 5′-GTCGTT-3′ and 5′-GACGTT-3(considered to be the most stimulatory for human and murine cells, respectively [24,27]) were determined using *compseq* (EMBOSS). For strains R-18194, R-1808 and LB400^T, we extrapolated the absolute number of CpG containing motifs based on the numbers observed in the contigs listed in Table 1 and the estimated total genome size. Statistical analyses were performed using the SPSS 11.0.1 software package (SPSS Inc., Chicago, IL).

3. Results and discussion

GC content, CpG dinucleotide relative abundance and absolute number and frequencies of CpG-containing motifs are shown in Table 1. The ρ^*_{CG} values of all *Burkholderia* genomes were significantly higher than those of the other genomes ($P<0.001$). The total number of CpG motifs ranged from 145,046 (*H. influenzae*) to 2,819,378 (*B. cepacia* complex R-18194). The lowest density of CpG

Table 1
GC content, CpG dinucleotide relative abundance and absolute number and frequencies of CpG-containing motifs

Species and strain designation	Genome size (bp)	GC content (mol%)	ρ^*_{CG}	CpG		5′-GTCGTT-3′		5′-GACGTT-3′	
				No.	Freq. (%)	No.	Freq. (%)	No.	Freq. (%)
Bordetella pertussis TohamaI	4,086,189	67.7	1.1505	1,078,033	13.19	1674	0.02048	1145	0.01401
Bradyrhizobium japonicum USDA110	9,105,828	64.1	1.2841	2,399,027	13.17	4804	0.02638	4992	0.02741
Burkholderia cenocepacia J2315*	7,963,121	67.0	1.4949	2,689,262	16.71	5573	0.03462	4768	0.02962
Burkholderia pseudomallei K96243	7,247,547	68.1	1.5502	2,602,216	17.95	4189	0.02890	4381	0.03022
Burkholderia mallei ATCC 23344	5,835,527	68.3	1.5521	2,123,594	18.19	3370	0.02887	3441	0.02948
Burkholderia cepacia complex R-18194 contig 88*	1,344,092	66.7	1.4875	440,640	16.39	974	0.03623	816	0.03036
Extrapolated for entire genome	±8600,000	–	–	2,819,378	–	6232	–	5221	–
Burkholderia vietnamiensis R-1808 contig 437*	411,398	65.7	1.4579	129,366	15.72	280	0.03403	260	0.03403
Extrapolated for entire genome	±8200,000	–	–	2,578,528	–	5581	–	5182	–
Burkholderia xenovorans LB400^T contig 715*	1,927,480	62.7	1.3778	522,738	13.56	1218	0.03160	1029	0.02669
Extrapolated for entire genome	±9700,000	–	–	2,630,667	–	6130	–	5178	–
B. cepacia complex SAR-1 scaffold 2204886	2,098,317	66.7	1.4761	687,966	16.39	1507	0.03591	1249	0.02976
Caulobacter crescentus CB15	4,016,947	67.2	1.1636	1,055,828	13.14	1990	0.02477	1735	0.02441
Chromobacterium violaceum ATCC 12472	4,751,080	64.8	1.1151	1,113,449	11.72	1221	0.01285	2076	0.02185
Desulfovibrio vulgaris Hildenborough	3,570,858	63.1	1.0706	762,107	10.67	2157	0.03020	1788	0.02504
Escherichia coli K12	4,639,221	51.8	1.1591	6,932,73	7.47	2314	0.02494	2970	0.03201
Haemophilus influenzae Rd	1,830,138	38.1	1.0897	145,046	3.96	677	0.01851	532	0.01453
Mycobacterium tuberculosis H37Rv	4,411,532	65.6	1.1832	1,123,596	12.74	2794	0.03167	2669	0.03025
Pseudomonas aeruginosa PAO1	6,264,403	66.6	1.1001	1,526,246	12.18	3106	0.02479	2258	0.01802
Ralstonia solanacearum GMI1000	5,810,922	67.0	1.1972	1,560,296	13.43	2852	0.02454	2239	0.01927
Staphylococcus aureus Mu50	2,878,040	32.9	0.9425	146,618	2.55	1192	0.02071	1155	0.02007
Xanthomonas campestris ATCC 33913	5,076,188	65.1	1.1323	1,216,768	11.99	2453	0.02416	1551	0.01528

All sequences were obtained from GenBank except *obtained from http://www.sanger.ac.uk/Projects/B_cenocepacia/ (*B. cenocepacia*) and http://genome.jgi-psf.org/microbial/ (*B. cepacia* complex R-18194, *B. vietnamiensis* and *B. xenovorans*).

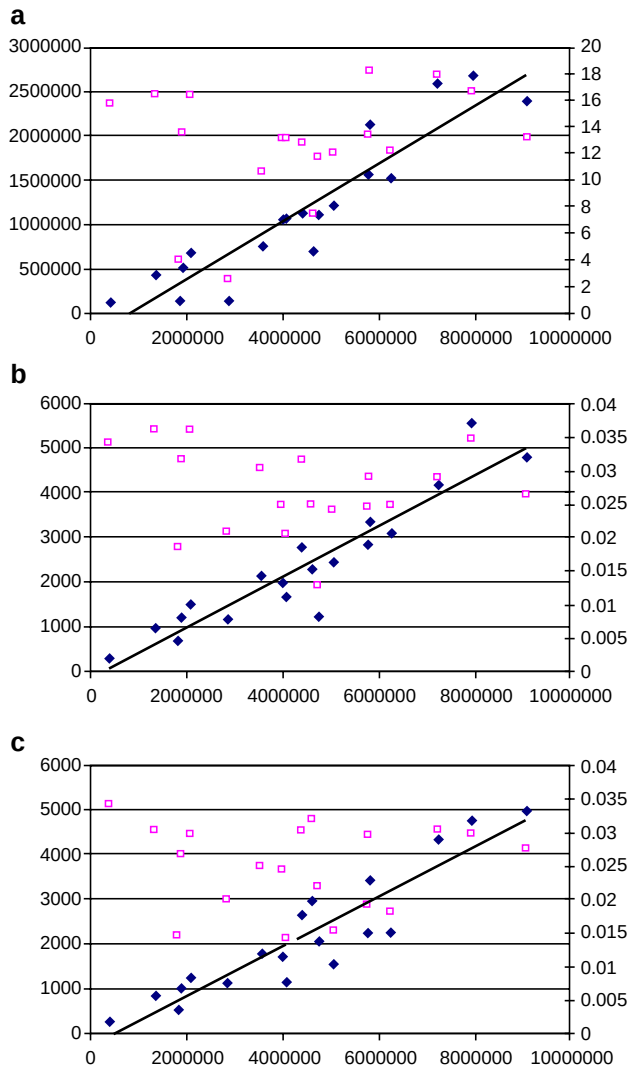

Fig. 1. Relation between genome size and absolute number and frequency of (a) CpG motifs, (b) 5′-GTCGTT-3′motifs and (c) 5′-GACGTT-3′motifs. *X*-axis: genome size (in base pairs). Left *Y*-axis: absolute number of motifs (solid diamonds). Right *Y*-axis: frequency (%) of motifs (open squares). The linear relation between genome size and absolute number of motifs are also shown.

motifs was found in the *S. aureus* genome (2.55%), while this density was highest in *B. mallei* ATCC 23344 (18.19%). The motif GTCGTT was most abundant in *B. cepacia* complex R-18194 (6232), while this motif was only encountered 677 times in the *H. influenzae* genome. The same trend was observed for the GACGTT motif, which was most abundant in *B. cepacia* complex R-18194 (5221), while this motif was only encountered 532 times in the *H. influenzae* genome. The density of the GTCGTT motif was also highest in *B. cepacia* complex R-18194 (0.03623%), while the *C. violaceum* genome had the lowest density for this motif (0.01285%). The highest GACGTT motif density was found in *B. vietnamiensis* (0.03403%), while the lowest density was found in the *B. pertussis* genome (0.01401%). Not unexpectedly, there is an almost perfect linear correlation between genome size and number of CpG motifs, GTCGTT motifs and GACGTT motifs ($r^2=0.925$, $r^2=0.932$ and $r^2=0.919$, respectively) ($P<0.01$ for all correlations) (Fig. 1). However, there is no correlation between genome size and motif density (Fig. 1). There is also a significant (but non-linear) relationship between GC content and the absolute number of the various CpG motifs (data not shown), as genomes with a higher GC content tend to have more of these repeats. However, when controlling for the effect of genome size (as larger genomes also tend to be more GC-rich), this correlation is not significant. When we compared the absolute number and frequency of CpG motifs, we found that the absolute number of CpG motifs is significantly higher in *Burkholderia* genomes than in non-*Burkholderia* genomes ($P<0.001$). The absolute number of GTCGTT and GACGTT motifs is also significantly higher in *Burkholderia* genomes than in non-*Burkholderia* genomes ($P<0.01$ for both motifs). In contrast to the absolute numbers, the differences in frequency are not statistically significant (data not shown).

Data from the present study show that *Burkholderia* genomes contain a large number of immunostimulatory CpG motifs and our statistical analyses indicate that this is mainly due to their large size. This suggests that DNA from *Burkholderia* sp. has the potential to cause significant inflammatory response. Whether this contributes significantly to the airway inflammation often observed in infected CF patients (either directly or by increasing the sensitivity to toxic effects of LPS and/or other bacterial products) remains to be determined.

Acknowledgements

T. C. and P. V. are indebted to the Fund for Scientific Research—Flanders (Belgium) for a position as postdoctoral fellow and research grants, respectively.

References

[1] Pilewski JM, Frizzell RA. Role of CFTR in airway disease. Physiol Rev 1999;79:S215–55.
[2] Gilligan PH. Microbiology of airway disease in patients with cystic fibrosis. Clin Microbiol Rev 1991;4:35–51.
[3] Bals R, Weiner DJ, Wilson JM. The innate immune system in cystic fibrosis lung disease. J Clin Invest 1999;103:303–7.
[4] Elborn JS, Shale DJ. Cystic fibrosis: 2. Lung injury in cystic fibrosis. Thorax 1990;45:970–3.
[5] Kronborg G. Lipopolysaccharide (LPS), LPS-immune complexes and cytokines as inducers of pulmonary inflammation in patients with cystic fibrosis and chronic *Pseudomonas aeruginosa* lung infection. APMIS Suppl 1995;50:1–30.
[6] Berger M. Inflammatory mediators in cystic fibrosis lung disease. Allergy Asthma Proc 2002;23:19–25.
[7] Wieland CW, Siegmund B, Senaldi G, Vasil ML, Dinarello CA, Fantuzzi G. Pulmonary inflammation induced by *Pseudomonas aeruginosa* lipopolysaccharide, phospholipase C, and exotoxin A:

role of interferon regulatory factor 1. Infect Immun 2002;70: 1352–8.
[8] Hughes JE, Stewart J, Barclay GR, Govan JR. Priming of neutrophil respiratory burst activity by lipopolysaccharide from *Burkholderia cepacia*. Infect Immun 1997;65:4281–7.
[9] Shaw D, Poxton IR, Govan JR. Biological activity of *Burkholderia* (*Pseudomonas*) *cepacia* lipopolysaccharide. FEMS Immunol Med Microbiol 1995;11:99–106.
[10] Zughaier SM, Ryley HC, Jackson SK. Lipopolysaccharide (LPS) from *Burkholderia cepacia* is more active than LPS from *Pseudomonas aeruginosa* and *Stenotrophomonas maltophilia* in stimulating tumor necrosis factor alpha from human monocytes. Infect Immun 1999;67:1505–7.
[11] Lever MS, Nelson M, Ireland PI, Stagg AJ, Beedham RJ, Hall GA, et al. Experimental aerogenic *Burkholderia mallei* (glanders) infection in the *BALB/c* mouse. J Med Microbiol 2003 (Dec);52 (Pt 12):1109–15.
[12] Friedland JS, Suputtamongkol Y, Remick DG, Chaowagul W, Strieter RM, Kunkel SL, et al. Prolonged elevation of interleukin-8 and interleukin-6 concentrations in plasma and of leukocyte interleukin-8 mRNA levels during septicemic and localized *Pseudomonas pseudomallei* infection. Infect Immun 1992;60:2402–8.
[13] Tokunaga T, Yamamoto H, Shimada S, Abe H, Fukuda T, Fujisawa Y, et al. Antitumor activity of deoxyribonucleic acid fraction from *Mycobacterium bovis* BCG: I. Isolation, physicochemical characterization, and antitumor activity. J Natl Cancer Inst 1984;72:955–62.
[14] Gilkeson GS, Grudier JP, Karounos DG, Pisetsky DS. Induction of anti-double stranded DNA antibodies in normal mice by immunization with bacterial DNA. J Immunol 1989;142:1482–6.
[15] Messina JP, Gilkeson GS, Pisetsky DS. Stimulation of in vitro murine lymphocyte proliferation by bacterial DNA. J Immunol 1991;147:1759–64.
[16] Tokunaga T, Yano O, Kuramoto E, Kimura Y, Yamamoto T, Kataoka T, et al. Synthetic oligonucleotides with particular base sequences from the cDNA encoding proteins of *Mycobacterium bovis* BCG induce interferons and activate natural killer cells. Microbiol Immunol 1992;36:55–66.
[17] Yamamoto S, Yamamoto T, Shimada S, Kuramoto E, Yano O, Kataoka T, et al. DNA from bacteria, but not from vertebrates, induces interferons, activates natural killer cells and inhibits tumor growth. Microbiol Immunol 1992;36:983–97.
[18] Kuramoto E, Watanabe N, Iwata D, Yano O, Shimada S, Tokunaga T. Changes of host cell infiltration into Meth A fibrosarcoma tumor during the course of regression induced by injections of a BCG nucleic acid fraction. Int J Immunopharmacol 1992;14:773–82.
[19] Krieg AM, Yi AK, Matson S, Waldschmidt TJ, Bishop GA, Teasdale R, et al. CpG motifs in bacterial DNA trigger direct B-cell activation. Nature 1995;374:546–9.
[20] Klinman DM, Yi AK, Beaucage SL, Conover J, Krieg AM. CpG motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma. Proc Natl Acad Sci U S A 1996;93:2879–83.
[21] Schmidt U, Wagner H, Miethke T. CpG-DNA upregulates the major acute-phase proteins SAA and SAP. Cell Microbiol 1999;1:61–7.
[22] Hartmann G, Weiner GJ, Krieg AM. CpG DNA: a potent signal for growth, activation, and maturation of human dendritic cells. Proc Natl Acad Sci U S A 1999;96:9305–10.
[23] Hemmi H, Takeuchi O, Kawai T, Kaisho T, Sato S, Sanjo H, et al. A Toll-like receptor recognizes bacterial DNA. Nature 2000;408:740–5.
[24] Bauer S, Kirschning CJ, Hacker H, Redecke V, Hausmann S, Akira S, et al. Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. Proc Natl Acad Sci U S A 2001;98:9237–42.
[25] Chuang TH, Lee J, Kline L, Mathison JC, Ulevitch RJ. Toll-like receptor 9 mediates CpG-DNA signaling. J Leukoc Biol 2002;71: 538–44.
[26] Hartmann G, Krieg AM. Mechanism and function of a newly identified CpG DNA motif in human primary B cells. J Immunol 2000;164:944–53.
[27] Hartmann G, Weeratna RD, Ballas ZK, Payette P, Blackwell S, Suparto I, et al. Delineation of a CpG phosphorothioate oligodeoxynucleotide for activating primate immune responses in vitro and in vivo. J Immunol 2000;164:1617–24.
[28] Klinman DM, Yi AK, Beaucage SL, Conover J, Krieg AM. CpG motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma. Proc Natl Acad Sci U S A 1996;93:2879–83.
[29] Hartmann G, Krieg AM. CpG DNA and LPS induce distinct patterns of activation in human monocytes. Gene Ther 1999;6:893–903.
[30] Cardon LR, Burge C, Clayton DA, Karlin S. Pervasive CpG suppression in animal mitochondrial genomes. Proc Natl Acad Sci U S A 1994;91:3799–803.
[31] Karlin S, Campbell AM, Mrazek J. Comparative DNA analysis across diverse genomes. Annu Rev Genet 1998;32:185–225.
[32] Karlin S, Mrazek J, Campbell AM. Compositional biases of bacterial genomes and evolutionary implications. J Bacteriol 1997; 179:3899–913.